ANALYSIS OF SEARCH METHODS OF OPTIMIZATION BASED ON POTENTIAL THEORY. I. NONLOCAL PROPERTIES

A. I. Kaplinskii, A. M. Pesin, and A. I. Propoi

UDC 517.977.5

Dynamic properties of nonlocal-optimization methods based on potential theory [1-3] are analyzed. The process of leaving the domain of a local minimum (the nonperspective domain) is considered. It is shown that the Newtonian potential generated by the objective function can be used as a Liapunov function of this search process.

1. INTRODUCTION

In [1-3] we studied methods of nonlocal optimization of nonsmooth functions based on potential theory. We considered methods of first and second orders (analogs of the gradient descent and the Newton method with respect to a potential function) that require only the values of the objective function for their implementation. Analyzing the properties of the potential function, we revealed the major structural components of nonlocal search: the unstable component in the nonperspective domain and the operations of reflection and dilatation of space in the nonperspective and perspective directions of search, respectively.

In this work, which is a continuation of [4], we study the convergence of these methods. We show that the Newtonian potential generated by the objective function and serving as a model of its extremal properties can be used as a Liapunov function of the search process.

A nonlocal search consists of two basic stages (processes) — the stage of leaving the nonperspective domain by the search point and the stage of localization of the solution (in the perspective domain).

In the first part of the work we analyze the problems of leaving of a domain by the trajectories of the search process.

2. STATEMENT OF THE PROBLEM

We consider the problem of unconstrained optimization

$$f(x) \to \min,$$

where the function $f : R^n \to R$. This problem is replaced by a randomized problem equivalent to the original problem [1]:

$$F(X) = E[f(X)] \to \min,$$

where E is the symbol of mathematical expectation, X is a random vector with values in R^n.

To find a solution of the optimization problem we use a random process in discrete time $N = 0, 1, \ldots$:

$$X^{N+1} = X^N + \alpha_N Y^{N+1}, \quad E\|X^0\| < \infty, E\|Y^1\| < \infty, \tag{1}$$

where the random vector Y^{N+1} defines the direction of search at the step N, α_N is a sequence of positive scalars.

Voronezh State University. Institute for System Analysis, Russian Academy of Sciences, Moscow. Translated from Avtomatika i Telemekhanika, No. 9, pp. 97–105, September, 1994. Original article submitted September 23, 1993.

The derivative of the functional $F(X)$ with respect to the direction Y has the form [1]

$$\delta_Y F(X) = -\int_{R^n} (f(x) - c)\,\mathrm{div}[p_X(x)\bar{y}(x)]\,dx,$$

where $p_X(x)$ is the probability density function of the random vector X, $\bar{y}(x) = E[Y \mid X = x]$, and c is an arbitrary constant.

It follows from these formulas that we can choose as a direction of search in (1) random vectors Y that satisfy the relations

$$\begin{aligned} &\mathrm{div}[p_X(x)\bar{y}(x)] \geqslant 0, \quad x \in D_N(c) = \{x \mid f(x) > c\}, \\ &\mathrm{div}[p_X(x)\bar{y}(x)] \leqslant 0, \quad x \in D_P(c) = \{x \mid f(x) \leqslant c\}, \end{aligned}$$

where the sets $D_N(c)$ and $D_P(c)$ define the nonperspective and perspective domains of the search, respectively.

As a particular solution of these inequalities we use the Newtonian field [5]:

$$p_X(x)\bar{y}(x) = \int_{D_N(c)} [f(s) - c]\,p_U(s)G(x, s)\,ds,$$

where

$$G(x, s) = \begin{cases} [(n-2)\omega_n \,\| x - s \|^{n-2}]^{-1}, & n > 2, \\ -\ln \| x - s \| \,(2\pi)^{-1}, & n = 2; \end{cases}$$

$$\|x - s\|^2 = \sum_{i=1}^{n} (x_i - s_i)^2,$$

ω_n is the area of the unit sphere in R^n, and $p_U(s)$ is a nonnegative function that can be considered as the density function of a random vector U.

The Newtonian field has the potential

$$\varphi(x) = -\int_{D_N(c)} [f(s) - c]\,p_U(s)G(x, s)\,ds.$$

In this case

$$\mathrm{div}[p_X(x)\bar{y}(x)] = \begin{cases} [f(x) - c]p_U(x), & x \in D_N(c), \\ 0, & x \notin D_N(c). \end{cases}$$

Thus, the process of nonlocal search is required to have the property of leaving of the nonperspective domain by the trajectories of (1). We find and discuss the corresponding conditions in the following sections of this work.

3. THE CONDITION OF LEAVING OF A DOMAIN BY THE TRAJECTORY OF SEARCH

In the following we shall assume that the minimum of the function f is attained in R^n. Let x^* be a (local) minimum of f. We introduce the sets

$$D^*(\varepsilon) = \{x \in R^n \mid K > f(x) > f(x^*) - \varepsilon\},$$

where

$$\infty > | K | > | f(x^*) - \varepsilon|, \quad 0 \leqslant \varepsilon < \infty.$$

For other points x of the search space R^n the corresponding domain is defined as

$$D(\varepsilon) = \{x \in R^n \mid K > f(x) - c > \varepsilon\},$$

where $\infty > K > \varepsilon > 0$ and the constant c is determined by the property of having fixed sign of the function $\widehat{f}(x) = f(x) - c$ [1-5].

We will use the function

$$\varphi(x) = \int_{D(\varepsilon)} \widehat{f}(s) p_U(s) G(x,s)\, ds. \tag{2}$$

as the Liapunov function of the process (1).

We denote by x a realization of the random vector X of the process (1). We assume that $x \in D(\varepsilon)$; moreover, because the set $D(\varepsilon)$ is an open set, we may assume that it includes x with a ball $B(x,r)$ of radius r and center at this point.

THEOREM 1. Let:

1) the vector process (1) have the Markov property, i.e.,

$$E[X^{N+1} \mid X^N, X^{N-1}, \dots, X^0] = E[X^{N+1} | X^N], \tag{3}$$

where $E\|X^0\| < \infty$, $E\|Y^1\| < \infty$;

2) the Liapunov function at the step N have the form

$$\varphi(N,x) = \int_{D(\varepsilon)} \widehat{f}(s) p_{U^N}(s) G(x,s)\, ds; \tag{4}$$

3) the random vectors Y^{N+1}, $N = 0,1,2,\dots$, that determine the direction of motion at the step N be normalized:

$$\int_{\|y\|=1} p_{Y^{N+1}}(y)\, dy = 1; \tag{5}$$

4) the distribition of the random vector U^N be concentrated in the ball $B(x, r^N)$, where x is a realization of the random vector X^N of the process (1), i.e.,

$$\int_{B(x,r^N)} p_{U^N}(s)\, ds = 1; \tag{6}$$

5) the conditional distribution of the random vector $X^{N+1} \mid X^N = x$ be concentrated in the domain $D(\varepsilon)$, i.e.,

$$\int_{D(\varepsilon)} p_{X^{N+1} \mid X^N = x}(s)\, ds = 1; \tag{7}$$

6) the density function p_{U^N} of the random vector U^N be limited within the ball $B(x, r^N)$:

$$p(r^N) \leqslant p_{U^N}(s) \leqslant P(r^N), \quad \infty > P(r^N) \geqslant p(r^N) > 0; \tag{8}$$

7) the scalar sequences $P(r^N)$, $p(r^N)$, α_N, r^N satisfy the conditions

$$\begin{gathered} \sum_{N=0}^{\infty} \alpha_N (r^N)^2 p(r^N) = \infty, \quad \sum_{N=0}^{\infty} \alpha_N = \infty, \\ \sum_{N=0}^{\infty} (\alpha_N)^2 r^N P(r^N) < \infty, \quad 0 \leqslant \alpha_N < \frac{\varepsilon}{4K} \frac{r^N p(r^N)}{P(r^N)}. \end{gathered} \tag{9}$$

Then the trajectories of the process (1) almost surely leave the domain $D(\varepsilon)$ at a finite time.

The proof of Theorem 1 is given in the Appendix.

4. DISCUSSION

The Markov property of the random process (1) (i.e., condition (3)) is a traditional assumption for the recurrent procedures of stochastic approximation type [6] (and process (1) is a procedure of this type). The new fact here is the nontraditional choice of the Liapunov function (4) with the "controlled" singularity of its kernel $G(x, s)$ (i.e., depending on the realization of the search point x). This leads to "pushing" of the search point out of the nonperspective domain [1]. Control of the singularity of the kernel of the Liapunov function is performed by the choice of the density fuction $p_U(s)$ satisfying conditions (8) and (9).

We give an example of the fulfillment of conditions (9) of Theorem 1 and show, in this way, that they are consistent.

For this we specify the sequence of scalars α_N, $N = 0, 1, \ldots$, by the condition $\alpha_N = O[(N+1)^{-\gamma}]$, $0 < \gamma < 1$. Obviously, the condition $\sum_{N=0}^{\infty} \alpha_N = \infty$ holds in this case.

If

$$(r^N)^2 p(r^N) = O[(N+1)^{-\kappa}], \quad 0 < \kappa < 1,$$

then the condition

$$\sum_{N=0}^{\infty} \alpha_N (r^N)^2 p(r^N) = \infty$$

will be satisfied, where the parameters γ and κ lie in the following boundaries:

$$0 < \kappa + \gamma < 1, \quad 0 < \kappa < 1/2 - \delta, \quad 1/2 > \delta > 0.$$

Here we may choose the boundaries of the parameters from the conditions

$$1/2 + \delta < \gamma < 1, \quad \delta < 1/2.$$

The condition

$$(r^N)^2 p(r^N) = O[(N+1)^{-\kappa}],$$

required in Theorem 1, may be fulfilled if

$$0 < r^N \leqslant R < \infty, \quad p(r^N) = O[(N+1)^{-\kappa}].$$

If

$$0 < P(r^N) \leqslant P < \infty, \quad r^N \leqslant R < \infty,$$
$$\alpha_N = O[(N+1)^{-\gamma}], \quad 1/2 + \delta < \gamma < 1,$$

we obtain

$$\sum_{N=0}^{\infty} (\alpha_N)^2 r^N P(r^N) < PR \sum_{N=0}^{\infty} O[(N+1)^{-2\gamma}] < \infty,$$

because $2\gamma > 1 + 2\delta$, $\delta > 0$.

In this case the inequality

$$0 \leqslant \alpha_N < \frac{\varepsilon}{4K} \frac{r^N p(r^N)}{P(r^N)}$$

takes the form

$$0 \leqslant O[(N+1)^{-\gamma}] < \frac{\varepsilon}{4K} \frac{r^N}{p(r^N)} O\,[(N+1)^{-\kappa}],$$

where $1/2 + \delta < \gamma < 1$, $1/2 > \delta > 0$.

If $P > P(r^N)$, then this inequality follows from the inequality

$$0 \leqslant O[(N+1)^{-\gamma}] < \frac{\varepsilon r^N}{4PK} O[(N+1)^{-\kappa}].$$

Assuming

$$p(r^N) = \mathrm{const}\,(N+1)^{-\kappa}, \quad \infty > \mathrm{const} > 0, \quad \alpha_N = (N+1)^{-\gamma},$$

we obtain from the latter inequality that

$$R \geqslant r^N > \frac{4PK}{\varepsilon} \mathrm{const}\,(N+1)^{-2\delta} \tag{10}$$

for all $N = 0, 1, 2, \ldots$.

As a result we obtain the following relations:

$$\begin{gathered}
\left[\frac{4PK}{\varepsilon} \mathrm{const}\right]^2 \sum_{N=0}^{\infty} (N+1)^{-(\gamma+4\delta+\kappa)} < \\
\sum_{N=0}^{\infty} \alpha_N (r^N)^2 p(r^N) < R^2 \sum_{N=0}^{\infty} [(N+1)^{-(\gamma+\kappa)}]\,\mathrm{const}; \\
0 < \sum_{N=0}^{\infty} (\alpha_N)^2 r^N P(r^N) < PR \sum_{N=0}^{\infty} (N+1)^{-2\gamma} < \infty, \quad \gamma > 1/2.
\end{gathered} \tag{11}$$

The inequality

$$0 \leqslant \alpha_N < \frac{\varepsilon}{4K} \frac{r^N p(r^N)}{P(r^N)}$$

follows now from (10).

In order to satisfy the system of inequalities

$$\begin{gathered}
0 < \gamma + \kappa < 1, \quad 1/2 + \delta < \gamma < 1, \\
0 < \kappa < 1/2 - \delta, \quad 0 < \delta < 1/2,
\end{gathered}$$

for $0 < \delta < [1 - (\gamma + \kappa)]/4 < 1/2$ it is sufficient to fulfill the inequality

$$\gamma + 4\delta + \kappa < 1.$$

This guarantees the fulfillment of the following condition:

$$\sum_{N=0}^{\infty} \alpha_N (r^N)^2 p(r^N) = \infty.$$

Finally, for $\delta = 1/16$ we can set $\gamma = 5/8$ and $\kappa = 1/8$ as an example of consistency of the inequalities for these parameters.

We note, in conclusion, that conditions (9) and (11) restrict the choice of the density function p_{U^N} and that they are compatible both with the constraint (8) and with the conditions for specification of the length step α_N required for leaving of the domain $D(\varepsilon)$ by the search trajectory.

We note also that conditions (9) will be satisfied if the limits

$$\lim r^N = 0, \quad \lim P(r^N) = 0, \quad N \to \infty$$

exist; in this case the limit

$$\lim r^N P(r^N) = \mathrm{const} < \infty, \quad N \to \infty.$$

must also exist.

This means the possibility of the choice of density fuctions that are δ-shaped and concentrated in balls of smaller and smaller radii if $N \to \infty$.

The above example establishes the limits of freedom for the choice of the parameters of process (1) satisfying all the conditions of Theorem 1. The choice of these parameters allows one to control effectively the nonlocal stage of the search (i.e., to control the process of leaving of the nonperspective domain by the search trajectory).

5. CONCLUSION

In this work, we obtained conditions that define the leaving of the nonperspective domain by the search trajectory. In these conditions we do not use the property of differentiability of the Liapunov fuction. An analysis of the process of nonlocal search taking into consideration the differential properties of the Liapunov function will be done in the next part of the work.

APPENDIX

Proof of Theorem 1. The proof is based on Theorem 5.1 from [6], which reads as follows.

Theorem [6]. Let $X(N)$ be a Markov random process in discrete time $N = 0, 1, \dots$ with generator L and an arbitrary initial distribution.

Let a nonnegative function $\varphi(N, x)$, where $N \leqslant 0, x \in D$, exist for which the estimate

$$L[\varphi(N, x)] \leqslant -q(N); \quad q(N) > 0, \ \sum_{N=0}^{\infty} q(N) = \infty$$

holds in a domain D.

Then the process $X(N)$ will leave almost surely the domain D at a finite time.

Here D is an open domain in R^m.

To prove Theorem 1 we shall show that all conditions of theorem 5.1 from [6] are fulfilled.

From condition (4) the inequality $\varphi(N, X) \geqslant 0$ follows, which is imposed in Theorem 5.1 from [6] upon the Liapunov function in the domain $D(\varepsilon)$.

We shall also prove the inequality

$$| E[\varphi(0, X^0)] | < \infty,$$

required in [6].

We introduce the following notations: $Y^{N+1} = Y$, $X^N = X$, $p_{U^N}(s) = p_U(s)$, $D(\varepsilon) = D$, $r^N = r$, $\alpha_N = \alpha$, $q_N = q$.

Making use of condition (3) and the Liapunov function (4), we define the increment of the generator L of the process (1) over the conditional measure $p_{X^{N+1}|X^N=x}(s)\, ds$ in the form

$$L[\varphi(N, x)] = \int_D p_{X^{N+1}|X=x}(s)[\varphi(N+1, s) - \varphi(N, x)]\, ds. \tag{A.1}$$

By using the conditions of Theorem 1, formula (1), and the definition of the domain D, we now evaluate the increment (A.1):

$$L[\varphi(N, x)] = \alpha \int_D p_{X^{N+1}|X=x}(x + \alpha y)[\varphi(N+1, x + \alpha y) - \varphi(N, x)]\, dy =$$

$$\alpha \int_D p_{X^{N+1}|X=x}(x + \alpha y) \left[\int_D \hat{f}(s) p_U(s) G(x + \alpha y, s)\, ds - - \int_D \hat{f}(s) p_U(s) E(x, s) \right] dy \leqslant$$

$$\alpha \int_D p_{X^{N+1}|X=x}(x + \alpha y) \left[K \int_D p_U(s) G(x + \alpha y, s)\, ds - \varepsilon \int_D p_U(s) G(x, s)\, ds \right] dy =$$

$$\alpha \int_D p_{X^{N+1}|X=x}(x+\alpha y)\left[K\int_{B(x,r)} G(x+\alpha y-s,s)p_U(s)\,ds-\varepsilon\int_{B(x,r)} G(x-s,s)p_U(s)\,ds)\right]\,dy \leqslant$$
$$\alpha \int_D p_{X^{N+1}|X=x}(x+\alpha y)\times\left[KP(r)\frac{2r\alpha}{(n-2)}-\varepsilon p(r)\frac{r^2}{2(n-2)}\right]\,dy =$$
$$\alpha\left[KP(r)\frac{2r\alpha}{(n-2)}-\varepsilon p(r)r^2\frac{1}{2(n-2)}2\right]=-q,\quad q>0. \tag{A.2}$$

From (9) it follows that

$$\sum_{N=0}^{\infty} q_N=\sum_{N=0}^{\infty}\alpha_N\left[-KP(r^N)\frac{2r^N\alpha_N}{(n-2)}+\varepsilon\,p\,(r^N)\frac{(r^N)^2}{2(n-2)}\right]=\infty.$$

Hence,

$$\sum_{N=0}^{\infty} q_N=\infty$$

which, by virtue of the conditions of Theorem 5.1 from [6], proves Theorem 1.

When deducing (A.2) we use the following estimates:

$$p(r)\int_{|R-\alpha|\leqslant r\leqslant R+\alpha} ds\frac{r^{2-n}}{\omega_n(n-2)}\leqslant I_1\leqslant P(r)\int_{|R-\alpha|\leqslant r\leqslant R+\alpha} ds\frac{r^{2-n}}{\omega_n(n-2)}; \tag{A.3}$$
$$I_1=\int G(x+\alpha y,s)p_U(s)\,ds=\int_D G(x+\alpha y-s,s)p_U(s)\,ds=$$
$$=\int_{B(x,r)} G(x+\alpha y-s,s)p_U(s)\,ds=\int_{|R-\alpha|\leqslant r\leqslant R+\alpha} p_U(s)\frac{r^{2-n}}{\omega_n(n-2)}\,ds;$$
$$p(R)\frac{2R\alpha}{(n-2)}\leqslant I_1\leqslant P(R)\frac{2R\alpha}{(n-2)}; \tag{A.4}$$
$$I_2=\int_D p_U(s)G(x,s)\,ds=\int_D p_U(s)G(x-s,s)\,ds=\frac{r^{2-n}}{\omega_n(n-2)}\int_{0\leqslant r\leqslant R} p_U(s)\,ds;$$
$$p(R)\frac{R^2}{2(n-2)}\leqslant I_2\leqslant P(R)\frac{R^2}{2(n-2)};\quad \mu=n-2. \tag{A.5}$$

For $\mu<n$ the relation

$$\int_{\hat r\leqslant r\leqslant\hat R} ds\,r^{-\mu}=\frac{\omega_n}{(n-\mu)}\left[\hat R^{n-\mu}-\hat r^{n-\mu}\right]$$

holds, which is obtained by the transition to the spherical system of coordinates [7].

Consequently,

$$\int_{\hat r\leqslant r\leqslant\hat R} G(r)\,dr=\frac{1}{2(n-2)}\left[\hat R^2-\hat r^2\right]. \tag{A.6}$$

The expression (A.6) is used in estimates (A.3), (A.4) and, respectively, in (A.2).

The conditions of Theorem 5.1 from [6] also include the inequality

$$|\,E[\varphi(0,X^0)]\,|<\infty.$$

We shall show that it is fulfilled here. In fact,

$$E[\varphi(0, X^0] = \int\limits_{R^n} p_{X^0}(x) \int\limits_D G(x, s)\hat{f}(s) p_U(s)\, ds$$

and, consequently,

$$|E(\varphi(0, X^0))| \leqslant K \int\limits_{R^n} p_{X^0}(x) \int\limits_D G(x, s) p_{U^0}(s)\, dx\, ds =$$

$$= K \int\limits_{R^n} p_{X^0}(x) \int\limits_{B(x,r)} G(x - s, s) p_{U^0}(s)\, dx\, ds \leqslant \frac{KP(r^0)(r^0)^2}{2(n-2)} < \infty,$$

because

$$P(r^0)(r^0)^2 < \infty, \quad K < \infty.$$

REFERENCES

1. A. I. Kaplinskii and A. I. Propoi, "Methods of nonlocal optimization based on potential theory," *Avtomat. Telemekh.*, No. 7, 55-64 (1993).
2. A. I. Kaplinskii and A. I. Propoi, "Nonlocal optimization methods of the first order based on potential theory," *Avtomat. Telemekh.*, No. 7, (1994).
3. A. I. Kaplinskii and A. I. Propoi, "Second-order conditions of improvement in nonlocal optimization methods based on potential theory," *Avtomat. Telemekh.*, No. 8, (1994).
4. A. I. Kaplinskii, A. M. Pesin, and A. I. Propoi, "On methods of nonlocal search," in: *Models and Methods of Optimization* [in Russian], VNIISI, Moscow, 35-45 (1991).
5. A. I. Kaplinskii and A. I. Propoi, *Structural Components of Nonlocal Search Methods Based on Potential Theory* [in Russian], VNIISI, Moscow (1990).
6. M. B. Nevel'son and R. Z. Has'minskii, *Stochastic Approximation and Recurrent Estimation* [in Russian], Nauka, Moscow (1972).
7. R. Courant and D. Gilbert, *Methods of Mathematical Physics*, Interscience, New York (1953).